

POSTER PRESENTATION

Open Access

SNP discovery in apple cultivars using next generation sequencing

Sérgio Alencar^{1*}, Orzenil Silva-Junior¹, Roberto Togawa¹, Marcos Costa¹, Luís Fernando Revers², Georgios Pappas¹

From IUFRO Tree Biotechnology Conference 2011: From Genomes to Integration and Delivery
Arraial d Ajuda, Bahia, Brazil. 26 June - 2 July 2011

Background

Knowledge about single nucleotide polymorphism (SNP) markers is extremely important in the development of genotyping assays, allowing improvements in plant breeding through marker-assisted selection. With the emergence of next generation sequencing platforms, high-density SNP discovery in the genome of plant crops becomes more achievable. In this project, we carried out whole genome resequencing of two apple cultivars (M13/91 and Fred Hough, developed by Epagri [1,2]) using the recently released draft genome sequence of the domesticated apple genome (*Malus x domestica* Borkh.) [3] as a reference for SNP discovery.

Materials and methods

We sequenced four DNA samples (two from M13/91 and two from FH cultivars) on four lanes of the Illumina GA II platform (single-end sequencing), with the Illumina PhiX sample used as a control. Image analysis and base calling were performed using on-instrument real time analysis, after which the Off-Line Basecaller was used to convert per-cycle base call files (.bcl) into per read base call files (.qseq). The resulting short reads were aligned to the reference genome with the BWA software [4], with the maximum edit distance set to 3, quality threshold for read trimming set to 15, and no gap opens allowed. Next, SAMtools [5] utilities were used to convert SAM into BAM format, to remove read duplicates and indels, and also to sort reads by coordinates. SNP discovery was then carried out using the GATK Unified Genotyper SNP caller [6].

Results

Alignment of the M13/91 and FH reads to the reference genome resulted in a total of 37,757,897 and 60,400,252 reads mapped, respectively. Considering only reads with root-mean-square mapping quality greater than 20, and occurring at raw read depths (DP) greater than 6 and lower than 20, a total of 143,468 and 474,483 heterozygous putative SNPs were identified when comparing the reference genome with the M13/91 and FH cultivars, respectively. A total of 80,554 heterozygous putative SNPs are shared by both M13/91 and FH cultivars. When considering only homozygous putative SNPs, a total of 20,296 (M13/91) and 70,659 (FH) SNPs were identified. A search was also made between the M13/91 and FH cultivar genomes, resulting in a total of 2,631 SNPs which are homozygous in FH and heterozygous in M13/91, and 4,768 SNPs which are homozygous in M13/91 and heterozygous in FH. In order to determine whether the differences in SNP frequencies between these cultivars are due to differences in read coverage obtained from sequencing, we set up a cut-off value above which all SNP calls in both cultivars had the same coverage, and it showed that their SNP frequency is similar.

Conclusions

We have used next generation sequencing data combined with high-density SNP detection methods to discover large numbers of putative SNPs in apple cultivars, which can be used in the development of genotyping assays.

Author details

¹Embrapa Recursos Genéticos e Biotecnologia - Parque Estação Biológica - Av. W5 Norte (final) Caixa Postal 02372 - Brasília, DF - 70770-917, Brazil.

²EMBRAPA Uva e Vinho - Rua Livramento, 515 - Caixa Postal 130 - Bento Gonçalves, RS - 95700-000, Brazil.

* Correspondence: salencar@cenargen.embrapa.br

¹Embrapa Recursos Genéticos e Biotecnologia - Parque Estação Biológica - Av. W5 Norte (final) Caixa Postal 02372 - Brasília, DF - 70770-917, Brazil
Full list of author information is available at the end of the article

Published: 13 September 2011

References

1. Denardi F, Hough LF: **Apple breeding in Brazil.** *Hortscience* 1987, **22**:1231-1233.
2. Denardi F, Camilo AP: **Novas cultivares de macieira: proposta de nova composição de pomares com polinizadoras/produtoras.** *Agropecuária Catarinense* 1997, **10**:25-30.
3. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchiotti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagné D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouzé P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel CE, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R: **The genome of the domesticated apple (*Malus x domestica* Borkh.).** *Nature Genetics* 2010, **42**:833-839.
4. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler Transform.** *Bioinformatics* 2009, **25**:1754-60.
5. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**(16):2078-79.
6. Depristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ: **A framework for variation discovery and genotyping using next-generation DNA sequencing data.** *Nature Genetics* 2011, **43**(5):491-498.

doi:10.1186/1753-6561-5-S7-P42

Cite this article as: Alencar et al.: SNP discovery in apple cultivars using next generation sequencing. *BMC Proceedings* 2011 **5**(Suppl 7):P42.

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

